CHIRAG LAKHANI

Website: https://cmlakhan.github.io

Email: chirag_lakhani@hms.harvard.edu

EDUCATION

North Carolina State University *December 2010*
Ph.D., Mathematics
Thesis: The GIT Compactification of Quintic Threefolds
Advisor: Amassa Fauntleroy
Designated emphasis on new algebraic geometry approaches to topological string theory; close collaboration with the algebraic geometry group at NCSU and the string theory group at Duke. Particular interest in toric varieties, mirror symmetry, and geometric invariant theory.

North Carolina State University *December 2003*
B.S. Mathematics
Minor: Physics
Magna Cum Laude

EXPERIENCE

Harvard Medical School - Training Program in Bioinformatics Applied to Diabetes, Obesity and Metabolism September 2018 - Present
T32 Training Fellow *Boston, MA*

· I am also currently funded by a NIH Training Grant that aims to train postdoctoral fellows with a quantitative background to solve problems in human metabolic health, such as obesity and Type 2 diabetes.
· Mentored by Dr. James Meigs and Dr. Chirag Patel

Harvard Medical School - Department of Biomedical Informatics March 2015 - Present
Postdoctoral Research Fellow *Boston, MA*

· Led development of "Exposome Data Warehouse," a database of aggregated publicly available environmental data (socioeconomic factors, pollution, and climate) that links to electronic heath records.
· Conducted an analysis where we repurposed insurance claims data to conduct the largest twin study in the United States in order to understand the role of genetics and the environment for 560 phenotypes. We leveraged "Exposome Data Warehouse" to also understand the contribution of socioeconomic status, climate, and pollution exposure to these phenotypes.

HelloWallet, Inc April 2014 - March 2015
Data Scientist *Washington, DC*

· Led initiative to prototype and productionize a machine learning based transaction classification system in the HelloWallet app. Work ranged from data cleaning, prototyping algorithm in sci-kit learn to develop technical approach for productionizing app in Java based app stack.
· Led data engineering initiative to migrate analytics database from MySQL environment to a Hadoop/Spark based environment.

Zaloni, Inc February 2013 - April 2014
Data Scientist *Research Triangle Park, NC*

· Led new data science initiative at Zaloni, implementing large-scale machine learning solutions to meet clients needs (such as fraud detection or scalable clustering algorithms for customer segmentation). Liaison between clients and development teams in the US and India to create data science products of value. Developed "Introduction to Data Science" training course focusing on applying Mahout to data science problems in the real world; taught at several Fortune 500 companies.

Duke University and SAMSI   August 2012 - February 2013
Independent Researcher   *Research Triangle Park, NC*

· Focused on developing scalable solutions to traditional nonlinear dimensionality reduction algorithms. Collaborated with Duke Prof. Mauro Maggioni and his group on leveraging geometric multi-resolution analysis to compute manifold learning algorithms efficiently.
· Primarily focused on SAMSI working group led by Prof. Ilse Ipsen (NC State), Research Scientist Michael Mahoney (Stanford), and Prof. Petros Drineas (RPI), on applying randomized linear algebra techniques to Kernel PCA algorithms for more efficient, scalable computing.

North Carolina State University   August 2011 - December 2011
Postdoctoral Research Associate   *Raleigh, NC*

· Advisor: Hamid Krim.
· Collaborated with Prof. Hamid Krim and his group on applying computational topology to problems in data mining. Primarily interested in the use of persistent cohomology for nonlinear dimensionality reduction. Also considered the use of Morse theory for detecting community structures in complex networks.

High Point University   August 2010 - May 2011
Instructor of Mathematics   *High Point, NC*

· Taught various math courses in the department such as calculus I, calculus II, and mathematics for elementary school teachers.

PUBLICATIONS

S. Cromer*, <u>C. Lakhani*</u>, D. Wexler, S. Burnett-Bowie, M. Udler, and C. Patel, "Geospatial Analysis of Individual and Community-Level Socioeconomic Factors Impacting SARS-CoV-2 Prevalence and Outcomes", *medRxiv* **medRxiv:2020.09.30.20201830** (2020)

Y. He, S. Groha, K. Taraszka, <u>C. Lakhani</u>, L. Braunstein, W. Foulkes, P. Polak, D. King, R. Tell, K. White, N. Zaitlen, C. Patel, and A. Gusev, "Genetic Ancestry and Population Differences in Somatic Alterations and Clinical Outcomes for Five Common Cancers", *In Review* (2020)

Y. He, <u>C. Lakhani</u>, A. Manrai, and C. Patel, "Poly-Exposure and Poly-Genomic Scores Implicate Prominent Roles of Non-Genetic and Demographic Factors in Four Common Diseases in the UK", *bioRxiv* **bioRxiv:833632** (2020)

<u>C. Lakhani*</u>, B.Tierney, A. Manrai, J. Yiang, P.M. Visscher, and C. Patel, "Repurposing large health insurance claims data to estimate genetic and environmental contributions in 560 phenotypes", *Nature Genetics* **51, 327 - 334** (2019)

<u>C. Lakhani*</u>, "The GIT Compactification of Quintic Threefolds", *arXiv* **arXiv:1010.3803** (2010)

* denotes first author

BIOMEDICAL TRAINING

Harvard Courses:

CS 281: Advanced Machine Learning	Fall 2015
EPI 249: Molecular Biology for Epidimiologists	Fall 2016
BST 227: Introduction to Statistical Genetics	Fall 2016
EPI 511: Advanced Population and Medical Genetics	Spring 2017

Summer Institute in Statistical Genetics:

Association Mapping: GWAS and Sequencing Data	July 2017
Mixed Models in Quantitative Genetics	July 2017
Advanced Quantitative Genetics	July 2017
Statistical and Quantitative Genetics of Disease	July 2017

Workshops:

Workshop on Transcriptomics	September 2017
Introduction to Single-Cell RNA-seq	July 2019

ACADEMIC TALKS

"Building a Search Engine to Find Environmental and Phenotypic Factors Associated with Health and Disease" NSF Big Data Spokes Meeting - Washington, DC, March 2017.

"Building an Exposome API" Environmental Statistics Seminar - Harvard School Of Public Health, February 2016.

"Systematic and large-scale investigation of twin and sibling concordance of 1723 traits in a nationally representative health claims cohort" American Society of Human Genetics Platform Talk - Baltimore, MD, October 2015.

"Approximation of Kernel Matrices" SAMSI Working Group - SAMSI, November 2012.

"Tensor Decompositions in Signal Processing" Tensor Seminar - North Carolina State University, October 2011.

"Topological Data Analysis and Zig-Zag Persistence" Geometric Methods in Signal Processing - North Carolina State University, August 2011.

"Secondary Fan for Toric Varieties" Toric Varieties Working Seminar - Duke University, November 2009.

"Schemes and Varieties" Visiting Student Seminar - Tata Institute of Fundamental Research, July 2005.

COMPUTATIONAL SKILLS

Programming Languages	C/C++, Python, Java
Hadoop Tools	Map-Reduce Programming, Hive, Pig, HBase, Spark
Databases	Hive, MySQL, PostgreSQL, HAWQ, Impala, Microsoft SQL Server
Large Scale Machine Learning	Mahout, Graphlab, MADLib, MLlib
Machine Learning Libraries	Various R libraries, sci-kit learn, MATLAB, PyMC, STAN PyTorch, Pyro, JAX, GPyTorch

ONLINE PROFILES

Github	https://github.com/cmlakhan
Linkedin	https://www.linkedin.com/in/chiraglakhani
Website	http://cmlakhan.github.io/

HONORS AND AWARDS

Stellar Abstract Award - Harvard PQG Conference	November 2018
Cloud Computing Credits Award ($100,000) (Microsoft Azure)	May 2017
AMS Travel Grant - MRC (Computational and Applied Toplogy)	June 2011
NSF Travel Grant - MEGA 2011 Conference, Stockholm, Sweden	May 2011
NSF Travel Grant - Aspects of Moduli Conference, Pisa, Italy	June 2008
Department Undergraduate Award for Research	December 2003
PAMS College Undergraduate Award for Service	December 2003